



An Evaluation Roadmap for a More Effective Government

September 2019

About the American Evaluation Association

The American Evaluation Association is a professional association of evaluators devoted to the application and exploration of approaches to evaluate programs, policies, and activities. Evaluation involves assessing the strengths and weaknesses of programs, policies, personnel, products, and organizations to improve their effectiveness. The American Evaluation Association has over 7,000 members representing all 50 states in the United States and more than 80 countries.

Purpose of the Evaluation Roadmap

Consistent with the American Evaluation Association's mission, this document describes our vision of the role for evaluation in the federal government. This Evaluation Roadmap outlines steps to build government capacity for strengthening the practice of evaluation throughout program lifecycles and for decision-makers to use evaluation to inform policymaking.

The Evaluation Roadmap was first developed a decade ago to share the lessons learned in agencies that have applied evaluation. It has been used by several agencies in shaping their evaluation policies. The Evaluation Roadmap offers a strategy to guide agencies and the Congress as they strive to meet expectations for increasingly producing and using evidence to inform policymaking.

An Evaluation Roadmap for a More Effective Government

Evaluation is an essential function of an effective government. Obtaining systematic evidence about government performance is a core responsibility for program management. Evaluation is a form of analysis that applies systematic research methods to address questions about a program, policy, or activity. An evaluation's findings and recommendations can suggest improvements in the effectiveness and efficiency of government services, assess which programs or policies are working and which are not, provide critical information needed for making difficult decisions, and enhance oversight and accountability of federal programs.

The Challenge: An Essential Need for Evidence

The U.S. government faces an array of challenges in national security, foreign aid, energy, the environment, health care, education, and the economy. Government decision-makers are responsible for identifying and understanding problems, and then developing solutions by designing programs, regulations, policies, activities, or initiatives to mitigate or resolve problems. To determine the merit, quality, appropriateness, and usefulness of government's interventions, credible systematic information is needed about the program or policy in question. This information includes knowledge about operations, what has been achieved, and at what cost. Such information is crucial if government officials are to ensure that the chosen interventions are working, taxpayers' money is spent wisely, and government is accountable to the public for producing results.

Since the mid-20th century, federal agencies have conducted many formal evaluations and applied their results to make reasoned program decisions. But, until recently, many of these evaluations were sporadic, applied inconsistently, and supported inadequately. The Government Performance and Results Act of 1993 (GPRA), reinforced by the GPRA Modernization Act of 2010, sought to address these challenges by establishing an outcome-oriented framework of strategic planning, performance management, and evaluation for federal agencies and programs. However, agency progress in implementing the GPRA framework has been slow.

There have been some successes in strengthening government's performance management systems based on the framework. For example, GPRA efforts encouraged some program evaluation studies which can usefully supplement ongoing performance monitoring by addressing more complex questions about the value of a program or policy, or the reasons why a program is or is not performing as expected. Still, implementation of the collective GPRA framework was less successful at motivating federal agencies to conduct evaluations and produce a feedback loop for performance management and operational decision-making.

Thus, while the practice of evaluation continues to grow rapidly within federal agencies, the capacity to effectively implement evaluation activities varies greatly across the federal government. In 2017, the U.S. Commission on Evidence-Based Policymaking's members unanimously concluded that government needs to improve its capacity to generate useful evidence and then use that information in decision-making.

Following the Commission's statement, the Foundations for Evidence-Based Policymaking Act of 2018 (P.L. 115-435) outlined an expectation that federal agencies establish a core

infrastructure to support and implement an evaluation function. The Evidence Act aims to increase both agencies' production and their use of a portfolio of evaluation and other systematic evidence. The Evidence Act's required infrastructure for agencies' evaluation activities now includes a senior responsible leader or evaluation officer, strategic planning for research and evaluation, written evaluation policies and procedures, support for workforce expertise to implement the function, and reporting on agency capabilities for analysis and evaluation.

In order to fully realize the potential benefits evaluation has to offer in pursuing a more effective government, federal agencies, executives, program managers, and other staff are now faced with the challenge of effectively leveraging new statutory requirements for building evaluation capacity.

The Opportunity: Using Evaluation for Effective Governance

Evaluation involves assessing the strengths and weaknesses of programs, policies, and organizations to improve their effectiveness, efficiency, and worth. It uses systematic data collection and analysis to address questions about how well government programs and policies are working, whether they are achieving their objectives, and, no less importantly, why they are or are not effective. Evaluation provides a useful and important tool to bring credible, well-grounded evidence to bear on a broad range of government decisions. It supports the development of new knowledge, innovation, and organizational learning in both the Legislative and Executive Branches.

Evaluation evidence can be used to compare alternative programs or policy options, guide program development and decision making, and reveal effective practices. The courts, as well, may cite evaluation findings as a basis for their judgments. More fundamentally, evaluation can contribute the evidence needed to support the system of checks and balances established by the U.S. Constitution. It enhances congressional oversight and executive accountability and, by its very nature, supplies publicly accessible information that is at the heart of transparency and open government. In addition, evaluation can address contextual appropriateness, providing evidence that can help ensure equitable practices and resources as well as appropriateness of programs across diverse settings.

Evaluation is an integral feature of good government, whether the goal is better performance, stronger oversight and accountability, or more data-informed and innovative decision making. The U.S. government will benefit significantly from using evaluation evidence in service of ends such as informing program and policy planning, monitoring program performance, and informing major decisions about program reform, expansion or termination. Other examples of evaluation's possible uses are listed in Exhibit 1.

The key to making the most that evaluation can offer is to make program evaluation integral to managing government programs at all stages, from planning and initial development through start-up, ongoing implementation, appropriations, and reauthorization. In short, what is needed is a transformation of federal management, budgeting, and policymaking to incorporate evaluation as an essential function.

Exhibit 1. Examples of Evaluation Possible Uses

- Address questions about current and emerging problems
- Articulate and test assumptions about the purpose and impact of policies or programs
- Provide timely feedback to decision makers enabling them to make changes when needed
- Increase government accountability and transparency
- Identify ways to reduce waste and enhance efficiency
- Improve programs and policies in a systematic manner
- Assess issues of equal access and equitable treatment across diverse communities
- Identify program implementation and outcome failures and successes
- Identify innovative solutions that work and in what contexts
- Inform the development of new programs where needed
- Examine the requirements for the transfer of promising programs to new sites or cultural contexts
- Share information about effective practices across government programs and agencies
- Re-examine program relevance and effectiveness over time.

Recommended Actions

We recommend that federal agencies adopt the following framework to guide the development and implementation of evaluation programs.

Scope and Coverage

- Conduct evaluations of public operations, programs, and policies throughout their life cycles and use evaluation both to improve programs and to assess their effectiveness
- Evaluate federal operations, programs, and policies in a manner that is appropriate for program stewardship and useful for decision making
- Build into each new program and major policy initiative an appropriate evaluation framework to guide and assess the program or initiative at key points throughout its life
- For existing programs, synthesize existing evaluations and other evidence to assess what is already known and develop evaluation plans to inform future decision making

Management

- Assign experienced, senior evaluation officials and managers to administer evaluation centers or coordinate evaluation functions at influential levels of government
- Prepare annual and long-term evaluation plans to guide decision making about programs
- Provide sufficient, stable funding to support evaluations and professional capacity building
- Ensure evaluation staff and contractors have the needed knowledge and skills, including technical competencies and cultural competence, for the scope of their work, according to recognized guiding principles and evaluator competencies from the American Evaluation Association
- Coordinate and communicate about evaluation efforts across agencies with overlapping or complementary missions

- Develop written evaluation policies across and within federal agencies that can guide evaluation efforts and help ensure their quality and appropriateness
- Ensure that evaluation units and staff receive high-level support

Quality and Independence

- Develop and adopt quality standards to guide evaluation functions consistent with the guiding principles for evaluators from the American Evaluation Association
- Promote the use and further development of appropriate methods for designing programs and policies, monitoring program performance, improving program operations, and assessing program effectiveness and cost
- Safeguard the independence of program or policy evaluations with respect to study design, conduct, results, and recommendations, while allowing for an appropriate level of consultation with and input from agency or staff personnel and other stakeholders
- Preserve and promote objectivity in examining program operations and impact

Transparency

- Consult closely with Congress and non-federal stakeholders in identifying program and policy objectives, critical operations, and definitions of success
- Communicate evaluation findings and methods to policymakers, program managers, and the public in contextually appropriate ways
- Create accessible information systems for publicly communicating and sharing evaluation findings about effective and ineffective program practices or policies

The next section fully develops these ideas. We describe the general principles that should guide a government-wide effort to implement the recommendations just presented, propose broad administrative steps to institutionalize evaluation in federal agencies, and discuss how the Executive Branch and Congress can jointly make the most effective and efficient use of evaluation.

Operational Principles Underlying Recommended Actions

The following general principles elaborate on the actions recommended above:

Scope

In general, all federal operations, programs, and policies should be subject to some sort of evaluation, although the appropriate evaluation approach will vary with the context and the available resources. Agencies should require periodic evaluations of each program throughout its life to provide rich evaluative information to policy makers during annual appropriation and authorization discussions.

Evaluation should be integral to planning, developing, managing, and implementing government programs at all stages of the program's life, including

- Ensuring that program designs are appropriate to achieve program goals
- Identifying problems during start-up and correcting them before they become entrenched

- Identifying and sharing promising approaches that emerge during program implementation
- Assessing the extent to which programs and policies are being well implemented
- To the extent feasible, and in consultation with stakeholders, establishing performance expectations at program inception and refining them as the programs mature
- Developing data collection and reporting procedures for performance measurement systems that provide a continuous flow of timely information to policymakers and program managers
- Examining the extent to which programs reach their intended beneficiaries and are culturally and contextually responsive
- Examining selected program features periodically to improve their effectiveness and efficiency
- Assessing program results and service quality periodically
- Examining systematically whether an apparently successful program can be expanded to another setting before scaling it up

Evaluation Policy and Procedures

Each federal agency and its evaluation centers or coordinators (discussed below) should publish policies and procedures and adopt quality standards to guide evaluations within its purview. Such policies and procedures should identify the kinds of evaluations to be performed and the criteria and administrative steps for developing evaluation plans and setting priorities, selecting evaluation approaches and methods to use, consulting subject matter experts, ensuring evaluation product quality, publishing evaluation reports, ensuring independence of the evaluation function, using an appropriate mix of staff and outside consultants and contractors, including diverse representation of staff and stakeholders, focusing evaluation designs and contracts appropriately, and promoting the professional development of evaluation staff.

Independence

Consultation with stakeholders is generally desirable on such matters as the questions to be addressed in an evaluation and the reporting schedule for study results. In addition, the leadership of government agencies and their component organizations can, and in many cases should, play a role in establishing general evaluation agendas, budgets, schedules, and priorities. At the same time, the real and perceived independence of the evaluation office must be preserved. Especially for evaluations performed in the service of public accountability, the evaluation organization must retain control of the evaluation's questions, design and methods, and most importantly, its results and distribution. Without such independence, the credibility of evaluations is at risk. However, the degree of shared decision making, and in some cases program staff participation in the conduct of an evaluation, may be greater for evaluations aimed at the improvement of ongoing programs and certain kinds of capacity building.

Analytic Approaches and Methods

Evaluators draw on an extensive array of analytic approaches and methods including, but not limited to, those in Exhibit 2.

The choice of analytic approaches and methods depends on the questions addressed, the purpose of undertaking the evaluation, the kind of program evaluated, the circumstances under which the program is implemented, its implementation status, the budgetary resources available, when the evaluation results are needed, and the evaluation's intended audience.

Furthermore, regardless of the choice, all evaluation methods should be context-sensitive, have cultural relevance, and be methodologically sound and rigorous.

Exhibit 2. Examples of Analytic Approaches and Methods Used in Evaluation

- Case studies
- Surveys and interviews
- Quasi-experimental designs
- Experimental designs/Randomized control trials (RCTs)
- Needs assessments
- Cost-benefit and cost-effectiveness analyses
- Early implementation reviews
- Logic models and evaluability assessments
- Evaluation synthesis and meta-analysis
- Systems analysis
- Culturally and contextually responsive approaches

Often, several analytic methods are needed in a single evaluation. Programs and their environments are complex; no single analytic approach or method can decipher the complexities in the program environment. Furthermore, definitions of “success” may be contested. And as a program matures, the right method or combination of methods may change. Some evaluation approaches are particularly helpful in a program’s early developmental stages, whereas others are more suited to ongoing, fully implemented programs or to analysis of temporary programs upon their completion.

How the evaluation is to be used in decision-making also can influence which approach or combination of approaches is best. When information is needed quickly, studies that can use existing data or rapid data-collection methods make the most sense. On the other hand, when the dynamics of the program and the behavior of its beneficiaries must be understood, then more sophisticated long-term studies may make more sense.

Likewise, depending on the intended use, evaluation methods might be used in combination with performance measurement or cost-effectiveness approaches. For example, performance measurement provides important, timely, routine information to managers and policy makers about what has happened in programs. Indeed, sometimes this is all they need. Adding evaluation to the mix makes sense when the question is *why* any changes happened or what changes, if any, would be beneficial.

Not surprisingly, for high-stakes decisions, a portfolio of evaluation and other sources of evidence is often needed. Multiple studies with similar results strengthen confidence in the conclusions, and a body of information can yield answers to a variety of different questions.

Resources

Evaluation activities in organizations should be supported through stable, continuous funding sources as well as through special one-time funds for evaluation projects of interest to Executive Branch and congressional policy makers. The stable, continuous evaluation funds should be provided through appropriations or program fund set-asides. These methods can also be combined to support viable evaluation programs. Each program should be evaluated throughout its life to provide credible evaluation information to decision makers. Under situations of severe budgetary constraints, where limited resources are available for evaluation, agency funding should go first to the highest-priority needs, with careful sequencing and prioritization of

ongoing work in terms of program and agency information needs as well as the potential of evaluations to provide information and insights that can guide action.

Professional Competence

Evaluations should be performed by professionals with appropriate training and experience for the evaluation activity (such as understanding diverse stakeholder perspectives, developing the study's design, performing statistical or other analyses, critiquing evaluation results, and communicating results credibly to diverse audiences). In 2018, the American Evaluation Association approved a set of evaluation competencies for use by its membership. Among these are interpersonal skills of communication and facilitation and ethical responsibilities to respect the dignity of all persons affected by the policies being evaluated. Evaluation is an interdisciplinary field that encompasses many areas of expertise. Many evaluators have advanced degrees in economics, political science, applied social research, sociology, anthropology, psychology, policy analysis, statistics, and operations research and often work collaboratively with colleagues in allied fields. Federal agencies should ensure that the required diversity of disciplines, including necessary expertise in the subject area being evaluated, and cultural backgrounds, is appropriately represented in internal and independent evaluation teams.

Evaluation Plans

Each federal agency should require its major program components to prepare annual and multiyear evaluation plans, and update these plans annually. The planning should take into account the need for evaluation results to inform program budgeting; reauthorization; agency strategic plans; ongoing program and regulation development and management; and responses to critical issues concerning program or initiative effectiveness, efficiency, and waste. These plans should include an appropriate mix of short- and long-term studies to produce results of appropriate scope and rigor for short- or long-term policy or management decisions. To the extent practical, the plans should be developed in consultation with diverse program stakeholders.

Evaluation questions can spring up unexpectedly and urgently. For example, questions may arise in response to an emerging issue, a sudden need for information to formulate a presidential initiative, resolve a management problem, or to respond to congressional inquiries. Therefore, evaluation plans should allow for flexibility in scheduling evaluations.

Dissemination and Sharing of Evaluation Results

To support the generation of knowledge about effective action, the results of all evaluations related to public accountability should be made available publicly and in a timely manner (except where this is inconsistent with the Freedom of Information Act, Privacy Act, national security, or other relevant laws). These results should be easily accessible and usable through the Internet. Similarly, evaluations of promising and effective program practices should be systematically and broadly disseminated to potential beneficiaries and to potential evaluation users in federal agencies. Reports and briefings on evaluation findings and methods, as well as their limitations, should be tailored so that various stakeholder audiences—policymakers, program staff, and others with an interest on the program or policy—can understand, interpret, and assess the validity, reliability, and credibility of the work. Evaluation data and methods should also—to the extent that can feasibly and ethically be achieved with sufficient privacy and

confidentiality protections—be made accessible to enable secondary analysis and assure transparency.

Evaluation Follow-Up and Tracking

To refine and improve the relevance of evaluation findings to policy- and program-based questions, agencies should track and re-examine the use made of those findings over time. For example, was the evidence generated from evaluations used by policy makers or program managers? If so, how? If not, why not? Did the program and its evaluation contribute over the longer term to resolving the public problem addressed? What could or should have been done differently? Again, the findings of such tracking should be shared with professionals and the public, to the extent possible.

Strengthening Evaluation Capacity in Government

Significant progress has been made in establishing evaluation as an integral component of government program management. However, additional steps are needed before agencies are consistently using program evaluation and systematic analysis to improve program design, implementation, and effectiveness and sharing what works, does not work, and why. There are historical precedents to build on and some new steps worth taking.

From the 1950s to the 1970s some federal departments established evaluation offices and used a variety of funding mechanisms to provide resources and expertise. In the 1970s, new sources of independent evidence on program effectiveness were created through what is now known as the Government Accountability Office (GAO) and agency inspectors general; these units conduct audits, evaluations, and investigations, including recommendations to promote agency economy, efficiency, and effectiveness.

The Government Performance and Results Act (GPRA) of 1993, modified by the GPRA Modernization Act of 2010, was one of the first government-wide efforts to encourage use of evidence and evaluation. GPRA requires each agency to develop a strategic plan and mission statement, and to determine whether their programs achieve their goals and objectives. In practice, most government agencies have implemented GPRA by using performance indicators and measurement to determine whether they have reached a goal. GPRA was unsuccessful in spurring agencies to conduct new evaluations that might provide insight about whether and how programs reached their goals, why programs do or do not meet their goals and objectives, and how to improve programs.

And most recently, the Foundations for Evidence-Based Policymaking Act of 2018 extends and strengthens this outcome-oriented framework by focusing on building agency capacity to produce and use credible evidence and enabling improved data sharing and use with strict privacy safeguards and greater transparency. The law directs agencies to develop evidence-building plans, establish evaluation officers, produce written evaluation policies, periodically report on their capacity to engage in statistics, evaluation, and policy analyses, and use such evidence for day-to-day governance.

To build on this legacy and continue to improve the capabilities for evaluation in government, and the ensuing benefits, both the Executive Branch and Congress will need to refine the role of evaluation moving forward:

Executive Branch Role

Many Executive Branch agencies now have a legal obligation to develop evaluation infrastructure and facilitate the practice of evaluation. Historically, the expectation was inconsistently applied; the Foundations for Evidence-Based Policymaking Act outlines an expectation for well-developed and stable evaluation practices. Other agencies, to which some provisions of the law do not apply, can also use the direction and expectation in conjunction with GPRA to develop robust evaluation capabilities.

Even with the legal requirement, federal agencies and programs will have different evaluation needs, which affects the design and organization of supporting operations. No single best practice or strategy exists for organizing evaluation offices and functions. While the 24 largest federal agencies are directed by law to develop certain activities department-wide, including an evaluation officer, there are multiple strategies for envisioning what an effective or useful operation may look like.

Central Evaluation Leader and Office. One option is to develop a central evaluation center at an agency, headed by an evaluation officer with support staff, the financial capability, and staff expertise to support evaluation activities across the entire agency. Employing this approach would likely require direct funding for a central evaluation unit as well as budgetary transfer authorities, such as in place at the U.S. Department of Labor, to support central evaluation activities.

Evaluation Leadership in a Federated System. In a federated system, different programs in an agency establish evaluation expertise and representatives that work in collaboration with a central evaluation officer at the agency. Individual agencies coordinate evaluation expenditures and resources with a chief evaluation officer, but the bulk of expertise and resources are retained by the individual programs. The U.S. Department of Health and Human Services has essentially operated a federated system since the 1970s, with some leadership provided by the Assistant Secretary for Planning and Evaluation, but decision-making about evaluation activities conducted within individual operating divisions like the Centers for Disease Control and the Administration for Children and Families. These each maintain their own evaluation leaders and funding capabilities separate from the central evaluation unit.

Evaluation Leadership with Decentralized Units. A decentralized evaluation approach in an agency recognizes that a central evaluation leader offers nominal guidance to support evaluation activities dispersed throughout the agency. This type of an approach may be preferable in agencies with a vast array of different types of activities and programs, such as the Department of Commerce. Decentralized approaches can also be useful in ensuring that the evaluation function remains strong and present during leadership turnovers or unexpected changes in resource allocations.

Whatever model is chosen, evaluation office(s) must include the functions and possess the attributes described above under general principles. Any of these approaches could be used to plan, conduct, and procure evaluation.

Legislative Branch Role

Building on the existing structures and organizations that support decision-making in Congress, a stronger connection between evaluations and legislative activities can emerge. This can be done by building evaluation expectations into authorizing legislation, explicitly setting aside adequate resources for evaluation, and signaling the importance of evaluation activities to agency officials.

Authorizations. Program authorization and periodic reauthorization provide opportunities for Congress to establish frameworks for systematic evaluation of new and continuing programs. Congressional committees can, through authorizing legislation, provide guidance or mandates relevant to evaluation. Related activities could include:

- Clearly identifying the intended goals of desired outcomes in legislation
- Encouraging early implementation reviews to identify start-up problems in such areas as scheduling, contracting, and grant making and to correct them before they become more serious
- Requiring evaluation planning at the outset of program design
- Providing adequate authority to collect meaningful data relevant for program analysis, or the sharing of data between relevant government agencies and programs
- Establishing or expanding ongoing surveys or other data-collection mechanisms to become relevant sources of reliable data
- Monitoring relevant performance indicators
- Facilitating studies reviewing the efficiency of program management and the fidelity implementation
- Encouraging targeted studies assessing program effects and identifying why programs are or are not effective, including interim results when long-term outcomes are of interest
- Requiring evaluation about topics of interest to Congress in support of authorization activities

Appropriations. In the annual appropriations process, Congress should ensure that agencies have sufficient resources to meet evidentiary and evaluation needs of decision-makers in Congress as well as the Executive Branch. Related activities could include:

- Funding for evaluation activities, including strengthening evaluation capacity, where needed (in terms of personnel, professional development, etc.)
- Developing funding set-aside authorities to enable agencies to transfer funding across budget accounts for evaluation activities
- Establishing flexible funding approaches and incentive funds to encourage innovation in research and evaluation
- Providing procurement flexibilities for long-term evaluations that may extend beyond the statutory period of availability for appropriated funds
- Posing questions to agencies about evaluation activities, and the insights of completed evaluations, during budget and appropriations hearings
- Requiring evaluation about topics of interest to Congress in support of appropriation activities

Oversight. The routine role of congressional oversight could lead to countless suggestions about facilitating strong evaluation activities in government. Activities could include:

- Holding hearings or less formal meetings to discuss evaluation findings, capacity, infrastructure, plans, or learning agendas with agency evaluation officers
- Monitoring agency implementation of the evaluation function, including establishment of evaluation units, appointment of qualified evaluation leaders, and transparency about evaluation activities
- Asking questions to individuals seeking Senate confirmation about their support and leadership interest in promoting strong evaluation policy in the agency

- Requiring evaluation about topics of interest to Congress in support of oversight activities

Collaboration Between the Executive and Legislative Branches

The usefulness of evaluation results could be maximized if Congress and the Executive Branch jointly specify broad evaluation expectations and concerns in authorizing statutes, appropriations, and other activities. Such collaboration will not always be easy or even possible to achieve. Nevertheless, when possible, a partnership of this kind can help increase the benefits that evaluation provides.

Congress and agencies can jointly work together to identify evaluation priorities, funding needs, and capabilities. The process of developing learning agendas in agencies should include input about congressional informational needs, as well as other stakeholders. The allocation of resources to evaluations should result from a dialogue about respective priorities, acknowledging timelines and knowledge gaps.

Looking to the Future

The Foundations for Evidence-Based Policymaking Act requires some federal agencies to establish evaluation infrastructures and workforces, but these activities alone will be insufficient in guaranteeing the use of evaluation findings and recommendations. Continued attention must be given by political and career leadership to encourage evaluative activities and to ensure their success moving forward.

The U.S. government continues to address major challenges to improve lives, protect the planet, and create efficiencies. With adequate resources; more thoughtful and systematic integration of evaluation into the planning, management, and oversight of programs; and increased application of evaluation results to planning and decision making, we can improve the performance of today's programs and ensure that tomorrow's programs reap the lessons learned today. Institutionalizing evaluation can also help achieve a more accountable, open, and democratic system of governance for future generations.

Additional information about the American Evaluation Association is available at www.eval.org.